Application of the Molecular Pseudopotential Method for Modeling the Toxicity of Chemical compounds

VLADIMIR MUKHOMOROV
Physical Department
State Polytechnic University
St. Petersburg
RUSSIA

Abstract: - Statistical modeling of the relationship between the toxicity of a number of substituted benzo-2,1,3-thia- and selenadiazoles depending on the number of various substituents and their position in the benzene ring was performed. It has been statistically reliably established that the toxicity of the analyzed series of chemical compounds is closely related to the value of the molecular pseudopotential. It has been shown that the relationship between the toxicity of drugs correlates linearly with the molecular electronic factor, which characterizes the magnitude of the pseudopotential of the molecule.

Key-Words: - Toxicity, benzo-2,1,3-thiadiazoles, pseudopotential, statistical modeling, electron factor, trend, correlation

Received: May 19, 2021. Revised: March 8, 2022. Accepted: April 10, 2022. Published: May 5, 2022.

1 Introduction

The antifungal activity of benzo-2,1,3-thiadiazoles is known [1], which was revealed on various test objects. However, there is no single method for initial trials of antifungal drugs. This makes it difficult to identify a quantitative relationship between the chemical structure of compounds and their toxic effects. Therefore, the method of modeling the relationship between the bioactivity of drugs and their molecular structure retains its relevance in connection with the need to clarify the mechanism of their action, as well as to predict new highly active chemical compounds.

2 Problem Formulation

In this article the toxicity of a series of substituted benzo-2,1,3-thia- and selenediazoles obtained and tested under the same conditions will be analyzed by statistical methods [1] as well as a quantitative relation between the structure of this series of compounds and their toxic action will be established. This will allow us to statistically reliably identify the most significant molecular parameters of chemical compounds that are responsible for activation in the biosystem of drugs, as well as to make some assumptions about the mechanism of their toxic effects.

3 Problem Solution

The search for the connection between the molecular structure of a compound and its toxicity is based on the idea that the objects under study have some effective electrostatic molecular potential, which is approximated by pseudopotential. Toxicity (LD_{50} in units of mg/kg) [1], the studied series of compounds is given in Table 1. Thus, the task is to choose a model with the minimum number of independent parameters that explain the largest fraction of the error variance. Determination of the real molecular potential is associated with complex quantum chemical calculations, which greatly complicates the construction of a practically convenient model. At the same time, the pseudopotential makes it possible to reliably reproduce many properties of condensed media. For example, the model pseudopotential correctly reproduces the nature of external electron scattering at atomic potentials in solids. The fact is that the scattering of an electron on a pseudopotential, which is rather weak, occurs in the same way as on a true potential. For a molecule, the pseudopotential is determined by the sum of the model potentials of the atoms that form the molecule [2]. It is important to emphasize that the consequences arising from the pseudopotential theory are in good agreement with the known experimental data on electron scattering [3].

In order to identify the relationship between the toxic properties of chemical compounds and their molecular structure, a method is proposed that uses

the average number of electrons in the outer shell of atoms in a molecule per atom as a factor sign of a molecule:

$$Z = \sum_i n_i Z_i / N, \quad (1)$$

where n_i is the number of atoms of the *i*th type with the number of electrons in the outer electron shell Z_i. The summation is performed over all atoms in the molecule; $\Sigma n_i = N$ is the total number of atoms. Within the framework of the pseudopotential method, it was shown [4] that the model pseudopotential of positive core ions of the molecule is weakened compared to the Coulomb field of an isolated core ion due to screening by external (valence) electrons

$$V(r) = \begin{cases} -Z|e|/r, & r > R_M, \end{cases} \quad (2)$$

where Z is defined by equation (1); $f(r)$ and $F(r)$ are corrections [4] to the Coulomb potential, depending on the distance r between the molecule backbone and electron; e is the electron charge, R_M is the scattering center radius. It can be shown that the parameter Z, which characterizes the number of electrons (2) is a common factor for the pseudopotential [5]. The model molecular pseudopotential method assumes that only electrons in the outer (valence) shell of the scattering center are taken into account. It is well known that the chemical properties of molecules are determined by the electronic state of a relatively small group of external electrons. The properties of the other electrons of the atom, which are called frame electrons, have almost no effect on the physical and chemical processes in which the molecule participates. This approximation is sometimes called the "frozen core approximation". In this approximation, the outer electrons do not move in the real Hartree-Fock force field of the molecule, but in a much weaker pseudopotential field. In this case, the behavior of the external electrons is close to their behavior in the Coulomb electrostatic field, and the pseudopotential itself is mainly determined by the first term of the potential (2). The physical meaning of the model pseudopotential is to describe the field of the core center in a complex molecular system or in a solid. The parameter that determines the potential variations in molecules is the average number of valence electrons per atom in a molecule. This result will be used in further studies.

Molecular potential can affect the biological system by interfering with the mechanisms that regulate life processes and thereby determine the biological activity of chemical compounds. According to the pseudopotential model, the average number Z of electrons on the outer electron shells of atoms in a molecule is used as a general factor (1) and (2) characterizing the molecular potential.

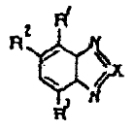

Fig.1. Molecular structure of substituted benzo-2,1,3-thia- and selenadiazoles. X = S or Se (see text for details).

Table 1
Toxicity [1] and molecular factor Z for substituted benzo-2,1,3-thia- and selenadiazoles.

N	Substitutes			$\log LD_{50}$	Z, arb. units
	R^1	R^2	R^3		
1	NO_2	Cl	H	1.80	4.40
2	NO_2	Br	H	1.69	5.07
3	NO_2	Cl	Cl	1.45	4.80
4	Cl	H	NO_2	1.81	4.40
5	Br	NO_2	NO_2	1.71	5.41
6	H	NO_2	H	2.60	4.00
7*)	NO_2	Br	H	1.45	5.73
8	OH	NO_2	NO_2	2.40	4.55
9*)	NO_2	H	H	1.71	4.67
10	NO_2	H	H	2.61	4.00
11	OC_2H_5	NO_2	NO_2	2.25	3.92
12	OC_4H_9	NO_2	NO_2	2.30	3.50
13	CH_2NH_2	H	H	2.78	3.24
14	COOH	H	H	2.78	3.75
15	OH	H	$COCH_3$	2.56	3.47
16	H	OC_2H_5	H	2.70	3.10
17	NH_2	NH_2	H	2.78	3.29
18	NH_2	CH_3	H	2.48	3.11
19	NH_2	H	H	2.30	3.33
20	H	NH_2	H	2.30	3.33
21	OH	H	H	1.88	3.57
22	H	OH	H	2.36	3.57
23	OH	CH_3	H	2.52	3.29
24	Cl	$OCH_2COOC_2H_5$	H	2.78	3.46
25	SH	H	H	2.00	3.57
26	SO_2H	H	H	2.70	3.88
27	$CH_2SPO(ONa)_2$	H	H	3.00	3.90
28	H	$CH_2NH(CH_2)_2SO(O)O(ONa)_2H$	H	2.78	3.38
29	$CH_2NH(CH_2)_2S_2O_3Na$	H	H	2.30	3.38

*) Selenium containing compounds.

This article analyzes the cause-and-effect relationship of toxicity - structure of the molecules

of a number of diazoles (Fig. 1). From the possible combination of different causes, an attempt will be made to identify the most significant, leaving aside secondary and incidental factors. Let us find out whether, for example, there is a trend between the toxicity of sulfur-containing compounds and the value of the explanatory molecular factor *Z*.

Let us find out whether, for example, there is a trend between the toxicity of sulfur-containing compounds and the value of the explanatory molecular factor *Z*. Let us construct a ranked series according to the toxicity value $\log LD_{50}$ for a sample composed of mutually independent elements. Each chemical compound of this series corresponds to a certain value of the molecular factor *Z* (Table 1). Let us determine whether the sequence of Z_i values for the ranked series by toxicity value is random, unrelated to the value of $\log LD_{50}$ or there is a systematic component to this sequence. To do this, we will use the Abbe-Linnik test [6,7]:

$$q = 0.5 \cdot \sum_{i=1}^{N-1} (Z_{i+1} - Z_i)^2 / \sum_{i=1}^{N} (Z_i - Z^{av})^2 =$$

$$0.305 < q_{0.05}^{cr}(N = 29) = 0.7047,$$

$$Z^{av} = N^{-1} \sum_{i=1}^{N} Z_i = 3.90,$$

$$Q^* = -(1-q)\cdot[(2N+1)/(2-(1-q)^2)]^{0.5}$$

$$= -4.33 < u_{0.05} = -1.645. \quad (3)$$

Since $q < q^{cr}$ and $Q^* < u_{0.05}$, the null-hypothesis of series randomness Z_i is rejected and the alternative hypothesis is accepted, which indicates a systematic shift of the mean [7] with confidence probability 0.95. It follows from inequalities (3) that the sequence of Z_i values has the following connection: the greater the value of the explanatory molecular factor Z_i, the more toxic the chemical compound. An approximate (for a sample size of less than 50) value of the parameter Q^* also indicates the existence of a systematic shift of the mean value. Using the toxicity data as well as the numerical values of the factor *Z* from Table 1, the corresponding scatter diagram can be constructed (Fig. 2). Scattering can be caused by unaccounted for factors (not necessarily secondary) or by chance, while the relationship becomes stochastic. Figure 2 demonstrates homoscedasticity - the relative stability and homogeneity of the random error variance of the regression model. Statistics (3) and the location of points (initial data) on the scatterplot suggests that there is a clear trend between the explanatory molecular factor *Z* and the resultant sign.

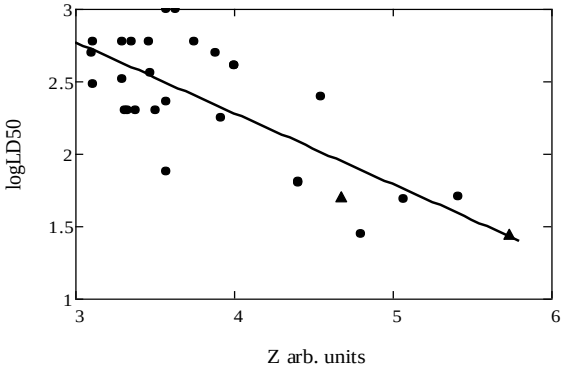

Fig.2. Scatterplot of toxicity observations ($\log LD_{50}$) for benzo-2,1,3-thia- and selenediazole derivatives. The regression line is determined by equation (4). • - sulfur-containing preparations, ▲ - selenium-containing preparations.

Indeed, using the methods of statistical analysis, it was found that between the explanatory sign *Z* and the toxicity ($\log LD_{50}$) of chemical compounds, there is the following averaged and statistically significant negative linear relationship:

$\log LD_{50}^{mod}(Z) = b_0^{(1)} + b_1^{(1)} \cdot Z$, $N = 29$, standard error of the regression estimate: $S_1 = 0.3125$; $R_1 = -0.73 \pm 0.09$; with a significant relationship between the explanatory variable and the resulting variable, the correlation coefficient should be significantly different from zero: $|R_1| > R_{0.05}^{cr}(f - m - 1) = 0.367$ [8]; sample size sufficient for the significance of the correlation coefficient: $N_{0.05}^{min} = 7$ [9]; $b_0^{(1)} = 4.10 \pm 0.33$, $b_1^{(1)} = -0.46 \pm 0.08$, $t(b_0^{(1)}) = 13.34 > |t(b_1^{(1)})| = 5.49 > t_{0.05}^{cr}(N-2) = 2.052$, $F = 30.15 > F_{0.05}^{cr}(f_1 = 1; f_2 = 27) = 4.21$; $\Delta = N^{-1} \cdot \Sigma_i(|\lg LD_{50i} - \lg LD_{50i}^{mod}| \cdot 100\% / \lg LD_{50i}) = 11\%$.

(4)

The significance of the regression coefficients b_i (4) is tested using Student's *t*-distribution (two-sided critical area) for $N - 2$ degrees of freedom and at a significance level of α. If $|t(b)| > t^{cr}$, then the regression coefficient *b* is significantly different from zero at the 95% confidence level. Since $F > F_{0.05}^{cr}(f_1; f_2)$, we can recognize the significance in general of the regression equation at $\alpha = 0.05$. The value of *F* is related to the coefficient of determination as follows [**10**]: $F = R^2(N-2)/(1-R^2)$. According to the Cheddock scale [12], if the

correlation coefficient falls into the range of values $0.7 < |R^*| < 0.9$, then the relationship between the variables is characterized as "strong (close) connection". For small samples, it is recommended [8] to use the corrected correlation coefficient: $R^* = R\cdot[1 + 0.5\cdot(1 - R^2)/(N - 3)]$. Using the data [9], you can specify the minimum sample size $N_{0.05}^{min} = 7 < N = 29$, sufficient for the reliability of the correlation coefficient $|R^*| = 0.74$ at the 95% confidence level.

Confidence limits of unit toxicity prediction $\log(LD_{50})$ for simple linear regression:

$$\pm t_\alpha(f = N - m - 1)S(Z). \quad (5)$$

Here $t_\alpha(f)$ is the quantile of the t-distribution with $f = N - m - 1$ degrees of freedom and significance level α; m is the number of explanatory variables. The value $S(Z)$ for the tested value Z can be calculated by the formula:

$$S(Z) = S_{ser}[1 + 1/N + (Z - Z^{av})^2/\sum_{i=1}^{N}(Z_i - Z)^2]^{0.5}, \quad (6)$$

here S_{ser} is the standard error of the residuals

$$S_{ser} = (N - m - 1)^{-0.5} \times \left[\left(\sum_{i=1}^{N}(\log LD_{50}^{mod}(Z_i) - (\log LD_{50})_i\right)^2\right]^{0.5}. \quad (7)$$

To determine a statistically significant linear correlation coefficient, it is necessary to fulfill the requirement of homogeneity of the analyzed sample [10]. It can be shown that the set of elements of $\log LD_{50}$ and Z (Table 1) are homogeneous and have a distribution close to the normal distribution at the 95% confidence level.

Population statistics of $\log LD_{50}$:

$N_1 = 29$, $\log LD_{50}^{av} = 2.34 \pm 0.09$; (2.16 - 2.51) is the confidence interval at significance level $\alpha = 0.05$; $\log LD_{50}^{min} = 1.45$, $\log LD_{50}^{max} = 3.00$; standard deviation: $S_{logLD1} = 0.46$; $\tau^{max} = 1.44 < \tau^{min} = 1.93 < \tau_{0.05}^{cr}(N_1) = 2.94$; Wilk-Shapiro normality test: $W = 0.918 \approx W_{0.05}^{cr}(N_1) = 0.928$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_1) = 3.47 \approx U = [(\log LD_{50}^{max} - \log LD_{50}^{min})/S_{logLD}] = 3.37 < U2_{0.05}^{cr}(N_1) = 4.89$; the coefficient of variation: $V = (19.66 \pm 2.58)\%$; representativeness of the sample size [9]: $N_{1repr} = 23$;

population statistics of the factor Z:

$N_1 = 29$, $Z^{av} = 3.90 \pm 0.13$; (3.63 - 4.17) is the confidence interval at significance level $\alpha = 0.05$; $Z^{min} = 3.10$, $Z^{max} = 5.73$, $S_{Z1} = 0.704$, $\tau^{min} = 1.14 < \tau^{max} = 2.60 < \tau_{0.05}^{cr}(N_1) = 2.94$; Wilk-Shapiro normality test: $W = 0.898 \approx W_{0.05}^{cr}(N_1) = 0.928$, David-Hartley-Pearson normality test: $U1_{0.05}^{cr}(N_1) = 3.47 < U = [(Z^{max} - Z^{min})/S_Z] = 3.73 < U2_{0.05}^{cr}(N_1) = 4.89$; $V = (18.05 \pm 2.37)\%$; $N_{1repr} = 23$. (8)

The homogeneity of the sample depends only on the sample size and is determined by the critical value of the Grubbs-Romansky τ-test [6,11]. If the volumes of samples Z and $\log LD_{50}$ are sufficient for the reliability of determining their main statistical indicators - mean values and standard deviations, then the volumes of these samples are also usually sufficient to identify the trend of the relationship between them [9].

The regression equation for the sample $N_2 = 27$, which includes only sulfur-containing drugs, practically does not differ from the regression (4):

$\log LD_{50}^{mod}(Z) = b_0^{(2)} + b_1^{(2)}\cdot Z$, $N_2 = 27$, $R = -0.65 \pm 0.12$, $|R^*| = 0.68 > R_{0.05}^{cr}(N_2 - 2) = 0.381$; sample size sufficient for the significance of the correlation coefficient: $N_{0.05}^{min} = 9$; $b_0^{(2)} = 4.02 \pm 0.36$, $b_1^{(2)} = -0.44 \pm 0.10$, $t(b_0^{(2)}) = 10.12 > |t(b_1^{(2)})| = 4.24 > t_{0.05}^{cr}(N_2 - 2) = 2.060$, $RMSE(S_2) = 0.321$; $F = 18.0 > F_{0.05}^{cr}(f_1 = 1; f_2 = 25) = 4.24$. (9)

The value of the empirical correlation ratio $\eta_{emp} = 0.679$ was determined for the relationship $\log LD_{50}^{mod}(Z)$ at $N_2 = 27$. Initial data after ranking by Z were divided into five groups: $n_1 = 5$, $n_2 = 5$, $n_3 = 6$, $n_4 = 6$, $n_5 = 5$. Intergroup variance $S_{lg}^2 = N^{-1}\cdot\Sigma_i(\log_i^{av} - \log^{av})^2\cdot n_i = 0.078$ and total variance $S_{lg}^2 = N^{-1}\cdot\Sigma_j(\log_j)^2 - (\log^{av})^2 = 0.169$, respectively. Here $\log_i^{av}$ is the values of the group means (options of the feature $\log LD_{50}$); $\log^{av}$ is the overall average value of the response function $\log LD_{50}$; index $i = 1,2,\ldots,5$; index $j = 1,2,\ldots,27$. In accordance with the Blackman curvilinearity criterion [9], we obtain the following inequality: $Bl = N_2\cdot|\eta_{emp}^2 - R^2| = 1.05 < Bl^{cr} = 11.37$, which indicates that the relationship between features should be straightline. An estimate of the curvilinearity of the relationship can also be obtained from the following relation:

$$t = 0.5\cdot N^{0.5}[(\eta_{emp}^2 - R^2)^{-1} - 2 + \eta_{emp}^2 + R^2]^{-0.5} = 0.523 < 3.00. \quad (10)$$

Since the inequality (10) is satisfied, we can agree that the analyzed relationship may slightly deviate from the straight-line dependence. The reliability of the correlation relationship is checked by the following relationship:

$$F = \eta^2_{\text{emp}}(N_2 - n)/(1 - \eta^2_{\text{emp}})/(n - 1) = 4.71 > F_{0.05}^{\text{cr}}(f_1 = n - 2;\ f_2 = N_2 - n) = 3.05. \quad (11)$$

A criterion based on Fisher's normalizing *z*-transformation [8]: $u = 0.825 > u_{0.05}(N_2) = z_{0.975}\cdot(N_2 - 3)^{-0.5} = 0.742$ also indicates the significance of the correlation coefficient. To check the difference between the coefficients $b_1^{(1)}$ (4) and $b_1^{(2)}$ (9), we use the following relationship [10]:

$$t = |b_1^{(1)} - b_1^{(1)}|\cdot \left(\frac{(N_1 - 2)S_1^2 + (N_2 - 2)S_2^2}{N_1 + N_2 - 4} \cdot \left(\frac{1}{(N_1 - 1)S_{Z1}^2} + \frac{1}{(N_2 - 1)S_{Z2}^2}\right)\right)^{-0.5}$$

$$= 0.216 < t_{0.05}^{\text{cr}}(N_1 + N_2 - 4) = 2.006, \quad (12)$$

which is valid because the following inequality holds (the ratio of the larger variance to the smaller variance): $F = (S_2/S_1)^2 = 1.051 < F_{0.05}^{\text{cr}}(f_1 = 25;\ f_2 = 27) = 1.93$. The inequality (12) quantitatively indicates the absence of a statistically significant difference between the regression coefficients $b_1^{(1)}$ and $b_1^{(2)}$. That is, the addition of selenium-containing compounds to a particular sample does not change the slope of the straightline regression (4). Additional information about the presence of a systematic shift in the average explanatory factor *Z* for the ranked toxicity of chemical compounds (e.g., only for sulfur-containing drugs; sample size $N = 27$) can be obtained by using the Abbe-Linnik test (3):

$$q = 0.2578 < q_{0.05}^{\text{cr}}(N = 27) = 0.6996,$$

$$Z^{\text{av}} = N^{-1}\sum_{i=1}^{N} Z_i = 3.80,$$

$$Q^* = -(1 - q)\cdot[(2N + 1)/(2 - (1 - q)^2)]^{0.5} = -4.73 < u_{0.05} = -1.645. \quad (13)$$

The smaller the empirical value *q* in comparison with the critical value, at the chosen level of significance, the clearer the relationship between the dependent feature and the explanatory variable. To make a statistical conclusion about the existence of a correlation, it is necessary to check the significance of the sample pair correlation coefficient. If there is a connection between the explanatory and resulting variables, the correlation coefficient *R* should be statistically significantly different from zero. The null hypothesis of no relation between the variables can be rejected if Fisher's inequality for *t*-distribution with $f = N - m - 1$ degrees of freedom is fulfilled for a sample correlation coefficient $R = -0.74$: $t_F = |R|\cdot(N - m - 1)^{0.5}/(1 - R^2)^{0.5} = 5.72 > t_{0.05}^{\text{cr}}(N - 2) = 2.052$. Here $m = 1$ is the number of explanatory variables. If the *t*-statistics calculated from the results of the sample is such that $t_F < t_{0.05}^{\text{cr}}(f)$, then a null hypothesis is accepted at a significance level of $\alpha = 0.05$, and the deviation of the correlation coefficient *R* from zero can only be attributed to unaccounted or random variations. For the studied relationship, the inequality $t_F > t_{0.05}^{\text{cr}}(f)$ was obtained, therefore, there is a significant statistical relationship between the variables. Note that a two-sided critical region is used here.

Let us use regression (4) to estimate the expected toxicity of a chemical compound that was not included in the original sample. Known [13] observed toxic dose $\log LD_{50} = 2.48$ for unsubstituted benzothiadiazole (gross formula $C_6H_4N_2S$; $Z = 3.39$ arb. units). Substituting the value $Z = 3.39$ arb. units into regression equation (4), we obtain the following toxicity estimate $\log LD_{50}^{\text{mod}} = 2.58$ for benzo-2,1,3-thiadiazole, which is close (comparison error < 4%) to the observed value of 2.48.

4 Conclusion

Checks were made on the use of additional explanatory variables in the regression equation. The Gammet constant σ_m of substituents in position R^1 of the benzene ring, the molar refraction $MR(R^1)$, which characterizes the volume size of substituents, as well as the contribution of π, which determines the hydrophobicity of substituents, were taken into account [14]. Taking these indicators into account did not improve the quality of the regression, and their contribution to the regression equation was statistically insignificant. However, it can be noted that those chemical compounds for which the substituents in the R^1, R^2, and R^3 positions preferably have a high electron affinity (i.e., they are electron acceptors) at the same time have the greatest toxicity. For example, replacing the hydrogen atom in position R^3 for the molecule (no. 1 in Table 1) with a chlorine atom (no. 3) significantly increases the toxic properties of the drug. A similar situation occurs when comparing molecule no. 5 with molecules nos. 8, 11, and 12. The electron affinity of the bromine atom in position R^1 is almost two times higher than that of the OH and OC_2H_5 substituents and 4–5 times higher than

that of the OC_4H_9 substituent. According to the energy of electron affinity, the substituents can be arranged in the following sequence: NO_2 > Cl > Br ≥ SH > OH > OC_2H_5 > NH_2 > OC_4H_9 > H [16].

As the explanatory factor *Z* increases, there is a tendency for the toxicity of chemical compounds to increase, and this tendency has a statistically significant linear character. Deviations from the regression line can be attributed to the influence of other unaccounted factors or random fluctuations.

Apparently, the molecular potential of benzo-2,1,3-thia- and selendiazole derivatives approximated by pseudopotential (1) - (2) determines the possible ability of chemical compounds to enter into paired intermolecular interaction with some region of the biophase and thereby initiate the toxic action of the drug. The greater the value of the molecular factor *Z*, the stronger the pairwise interaction of the molecule with the biophase region [16].

References:

[1] Belenkaya I.A., Dyachina Zh.S., Mukhomorov V.K., Sirik S.A., Structure-Activity. Relation of electronic parameters of substituted benzo-2,1,3-thia- and selenediazoles to their antifungal activity and toxicity, *Chemico-Pharmaceutical Journal*, No. 12, 1991, pp. 49-53.

[2] Abarenkov I.V., Bratsev V.F., Tulub A.I., *Beginnings of quantum chemistry*, Moscow, Higher school, 1989.

[3] Brandt N.B., Chudinov S.M., *Energy spectra of electrons and phonons in metals*, Moscow State University, Moscow, 1980.

[4] Veljkovič V., Lalovič D., Simple General – Model Pseudopotential, *Phys. Rev. Lett. A*, Vol.29, No2, 1972, pp.105-107.

[5] Husinaga S., *Method of molecular orbitals*, MIR, Moscow, 1983 (in Russian).

[6] Kobzar A.I., *Applied Mathematical Statistics. For engineers and* scientists, Moscow, 2016, (in Russian).

[7] Linnik Yu.V., *The Method of Least Squares and the Foundations of the Mathematical and Statistical Theory of Observation Processing*, Moscow, 1962, (in Russian).

[8] Sachs L,, *Statistische Auswertungsmetoden*, Berlin, 1972.

[9] Zaitsev G.N., *Mathematics in Experimental Botany*, Moscow, 1990, (in Russian).

[10] Förster E., Rönz B., *Metohden der Korrelations- und Regressionsanalyse*, Verlag Die Wirtschaft, Berlin, 1979.

[11] Dmitriev E.A., *Mathematical statistics in soil science*, Moscow State University Press, 1995, (in Russian).

[12] Baraz V.R., *Correlation-regression analysis of the relationship of commercial activity indicators using EXCEL programs*, State Educational Institution of Higher Professional Education "UGTU-UPI", Ekaterinburg, 2005.

[13] Belenkaya I.A., Slavachevskaya N.M., Strelnikov Yu.E., Prosypkina A.P., Synthesis of Substituted Benzo-2,1,3-thiadiazoles, *Chem. - Pharm. Journal.*, No.10, 1978, pp.25-30, (in Russian).

[14] Hansh C., Leo A., Unger S. H., Ki Hwan Kim, Nikaitani D., Lien E.J., "Aromatic" Substituent Constants for Structure-Activity Correlations, *J. Med. Chem.*, Vol.16, No.11, 1973, pp.1207-1216.

[15] Gurevich L.V., Karachevtsev G.V., Kondrat'ev V.N., Lebedev Yu.A., Medvedev V.A., Potapov V.K., Khodeev Yu.S., *The Energies of Breaking Chemical Bonds. Ionization Potentials and Electron Affinity*, Moscow. 1974. (in Russian).

[16] Mukhomorov V.K., Carcinogenic Activity of Polycyclic Hydrocarbons, *WSEAS Transaction on Biology and Biomedicine*, Vol.18, 2021, pp. 150-169.